

Systematic review of measurement properties of patient-reported outcome measures used in patients undergoing hip and knee arthroplasty

Kristina Harris¹
Jill Dawson²
Elizabeth Gibbons²
Chris R Lim¹
David J Beard¹
Raymond Fitzpatrick²
Andrew J Price¹

¹Nuffield Department of Orthopedics, Rheumatology and Musculoskeletal Sciences, ²Nuffield Department of Population Health (HSRU), University of Oxford, Oxford, UK

Objectives: To identify patient-reported outcome measures (PROMs) that have been developed and/or used with patients undergoing hip or knee replacement surgery and to provide a shortlist of the most promising generic and condition-specific instruments.

Methods: A systematic review of the literature was performed to identify measures used in patients undergoing hip and knee replacement and extract and evaluate information on their methodological quality.

Results: Thirty-two shortlisted measures were reviewed for the quality of their measurement properties. On the basis of the review criteria, the measures with most complete evidence to date are the Oxford Hip Score (OHS) (for patients undergoing hip replacement surgery) and the Oxford Knee Score (OKS), with OKS-Activity and Participation Questionnaire (for patients undergoing knee replacement surgery).

Conclusion: A large number of these instruments lack essential evidence of their measurement properties (eg, validity, reliability, and responsiveness) in specific populations of patients. Further research is required on almost all of the identified measures. The best-performing condition-specific PROMs were the OKS, OHS, and Western Ontario and McMaster Universities Osteoarthritis Index. The best-performing generic measure was the Short Form 12. Researchers can use the information presented in this review to inform further psychometric studies of the reviewed measures.

Keywords: knee, hip, patient-reported outcome measures, systematic review, measurement properties, validity, reliability.

Introduction

Patient-reported outcome measures (PROMs) are increasingly required to have adequate evidence of measurement properties following rigorous methodological standards.^{1–6} Whether used for research purposes (eg, in clinical trials), audit, or in clinical practice, PROMs should be chosen carefully after close inspection of evidence of their method of development, validity, reliability, responsiveness, context, and purpose.

Evolving standards of the acceptable evidence of measurement properties, coupled with selective reporting of such evidence in validation studies, somewhat complicate the choice of the most appropriate outcome measure for researchers and clinicians. For a researcher or a clinician who wants to choose the best outcome measure, keeping track of all the relevant literature sources can be challenging. Furthermore, repeated adjustment in the criteria of what is currently considered evidence of good measurement properties, coupled with the (mainly past) selective tolerance of “psychometric language” by some journals – so that not all relevant information has necessarily been

Correspondence: Kristina Harris
Nuffield Department of Orthopaedics,
Rheumatology, and Musculoskeletal
Sciences, Botnar Research Centre,
Windmill Road, Oxford, OX3 7LD, UK
Email: kristina.harris@ndorms.ox.ac.uk

published – makes the task of evaluating the best outcome measure for a specific purpose even more difficult.

The consequence of this lack of consensus about the minimum psychometric standards for a measure can be that researchers or clinicians can choose an unsuitable or a poor outcome measure for the intended purpose (eg, for a particular clinical trial). Results obtained using the inappropriate outcome measure are then incorporated into clinical decision-making models, which have the potential to influence patients' lives. Outside of the obvious scientific error, the issue introduces clear ethical concerns and implications.

The evolution of PROMs methodology and proliferation of various instruments over the years may have led to some confusion when choosing an appropriate measure for a data collection exercise (eg, audit/registry) or a research study.⁷ While the orthopedic literature is awash with different scoring systems and outcome measures specifically used for assessing the outcomes of hip or knee arthroplasty, not all measures have evidence of, or reach, even the minimum psychometric standards for their proposed uses. Indeed, to date, only a few measures for assessing the outcome of hip or knee arthroplasty have been shown to meet some of the criteria applied in reviews of their measurement properties.^{8–10}

The aims of this report are to identify and evaluate English-language versions of PROMs, which have been evaluated with patients undergoing hip or knee replacement surgery and to provide a comprehensive profile of their measurement properties.

Methods

Identification of studies

A sensitive filter for finding studies on measurement properties was used to search MEDLINE, EMBASE, PsycINFO, and AMED. ProQolid, Oxford PROMs Database, Dare, and Econlit were also searched using a combination of MeSH and free-text terms. The search was conducted in May 2014, was limited to English language, and no time restrictions were set. MEDLINE, EMBASE, PsycINFO, and AMED were searched using an adjusted methodological filter through OVID ([Supplementary material 1](#)).¹¹ Handsearching of titles of the following key journals in the 6 months preceding the search was also conducted: *Health and Quality of Life Outcomes*, *Journal of Bone and Joint Surgery* (Am and Br), and *Journal of Arthroplasty*.

Screening of articles and instruments

Titles and abstracts of all identified articles were assessed for inclusion/exclusion by two reviewers (KH and EG).

Agreement between reviewers was assessed on a test screening sample. Agreement was tested on a subsample of 313 abstracts. The first round of testing yielded a 77% agreement rate, and the second round yielded a 99% agreement rate, between the two reviewers. Full texts of the articles that are to be included in the review were retrieved. Inclusion criteria were

- The instrument uses a standard scoring system (representing indices or scales).
- The instrument is already available and has been used in clinical settings or research to assess adult (>18 years old) patients prior to hip or knee replacement.
- The instrument has been validated for the English-language population.
- The study design is principle development, concurrent revalidation, or a prospective study of a score with information on its measurement properties (eg, reliability, validity, and responsiveness). Retrospective studies (except historical cohort studies) were excluded.
- Sample size in the study was >50 subjects/patients.

Titles and abstracts were obtained relating to any tools identified at this stage, and these were scrutinized using the aforementioned inclusion criteria. Two members of the team conducted their respective tasks independently. The same methodology was applied on full-text documents for their inclusion in the review, as well as in the case of abstracts that were identified but where initial abstract-based information led to uncertainty or disagreement between assessors.

Selected full-text articles were then screened for all measures that were used in analyses. The aforementioned inclusion criteria were applied on the list of identified measures. Furthermore, the following exclusion criteria were applied on the initial list of measures.

- The assessment is not patient reported and requires the patient to be assessed on each/every occasion by a clinician.
- The assessment requires some kind of technical information or equipment (such as MRI scan or X-ray report), which might not always be available or standardized, or which might not make sense as part of an assessment conducted at both pre- and postoperative stages.
- The measure is not capable of demonstrating patients' "capacity to benefit" – because it was not designed to be a health status/outcome measure, and therefore cannot measure change, for example, purely retrospective measures were excluded.

Instrument-specific search

Additionally, a specific search was performed for each of the identified instruments. First, a developmental study was identified for each instrument. Then, a population and validation filter was applied ([Supplementary material 1](#)) to the list of citations stemming from the developmental study.

Data extraction

Data were extracted on the psychometric performance and operational characteristics of each PROM. Assessment and evaluation of the methodological quality of PROMs were performed independently by three reviewers adapting the London School of Hygiene appraisal criteria outlined in a previous review.¹⁰ These criteria were modified for our review ([Supplementary material 2](#)).

Reliability was assessed by looking at the test–retest reliability and internal consistency. Test–retest reliability refers to the stability of a measuring instrument over time; assessed by administering the instrument to respondents on two different occasions and examining the correlation between test and retest scores. Internal consistency refers to the extent to which items comprising a scale measure the same construct (eg, homogeneity of items in a scale) and is assessed by Cronbach's α and item-total correlations.

Validity was assessed by examining the content and construct validity. Content validity relates to the extent to which the content of a scale is representative of the conceptual domain it is intended to cover and is usually assessed qualitatively during the questionnaire development phase through pretesting with patients, with patients involved in item generation. Construct validity looks at the evidence that the scale is correlated with other measures of the same or similar constructs in the hypothesized direction and is assessed on the basis of correlations between the measure and other similar measures, preferably based on a priori hypothesis with predicted strength of correlation.

Responsiveness refers to the ability of a scale to detect significant change over time and is assessed by comparing scores before and after an intervention of known efficacy or where other evidence indicates important change on the basis of various methods including paired *t*-tests, effect sizes, standardized response mean values, or responsiveness statistics. Ideally evidence of responsiveness will include high correlations between the change scores of the scale and relevant constructs preferably based on a priori hypothesis with predicted strength of correlation.

Interpretability relates to the degree to which one can assign qualitative meaning – that is, clinical or commonly

understood connotations – to an instrument's quantitative change in score. It can be assessed by estimating the precision of the measure when used at an individual patient level, by multiplying standard error of measurement with the standard score (*z*-value). In addition, minimal clinically important differences/changes can be calculated by relating change to an external anchor, either using mean change or receiver-operating characteristic curve method.

Floor/ceiling effects relate to the ability of an instrument to measure accurately across full spectrum of a construct. If a measure has >15% of participants achieving top or bottom score, this is indicative of a ceiling/floor effect.

Acceptability is a practical property of an instrument and reflects respondents' willingness to complete it without feeling unduly burdened, indicated by, for example, response rates and completion rates.

Measurement properties for each instrument were assessed separately for hip, knee, or mixed hip/knee populations (depending on the availability of published studies). The information was then summarized into the appraisal summary tables, which rated the overall quality of evidence for each of the measurement properties. The scoring for each property is presented in [Supplementary material 2](#). Three authors (KH, EG, and JD) reviewed their own respective sections, following which the results were cross-checked to ensure consistency of assessment and scoring across the reviewers.

Results

Identification of studies

The initial search in OVID yielded 3,774 abstracts. After removal of duplicates, the number of abstracts for assessment was 2,887 (Figure 1). Additionally keyword searches (combination of knee, hip, and orthopedics) in EconLit yielded 162 results, PROMs Database identified 454 results, and Dare had no results.

Handsearching of titles of the following key journals in the 6 months preceding the search was conducted:

- *Health and Quality of Life Outcomes* (1),
- *Journal of Bone and Joint Surgery* (Am and Br) (1), and
- *Journal of Arthroplasty* (3).

Screening of articles and instruments

Out of 167 selected abstracts, ten abstracts were conference proceedings without full text, ten papers could not be found within the Oxford University Libraries, and one abstract was a book. One hundred and forty-six full-text articles were then screened for all PROMs that were analyzed. One

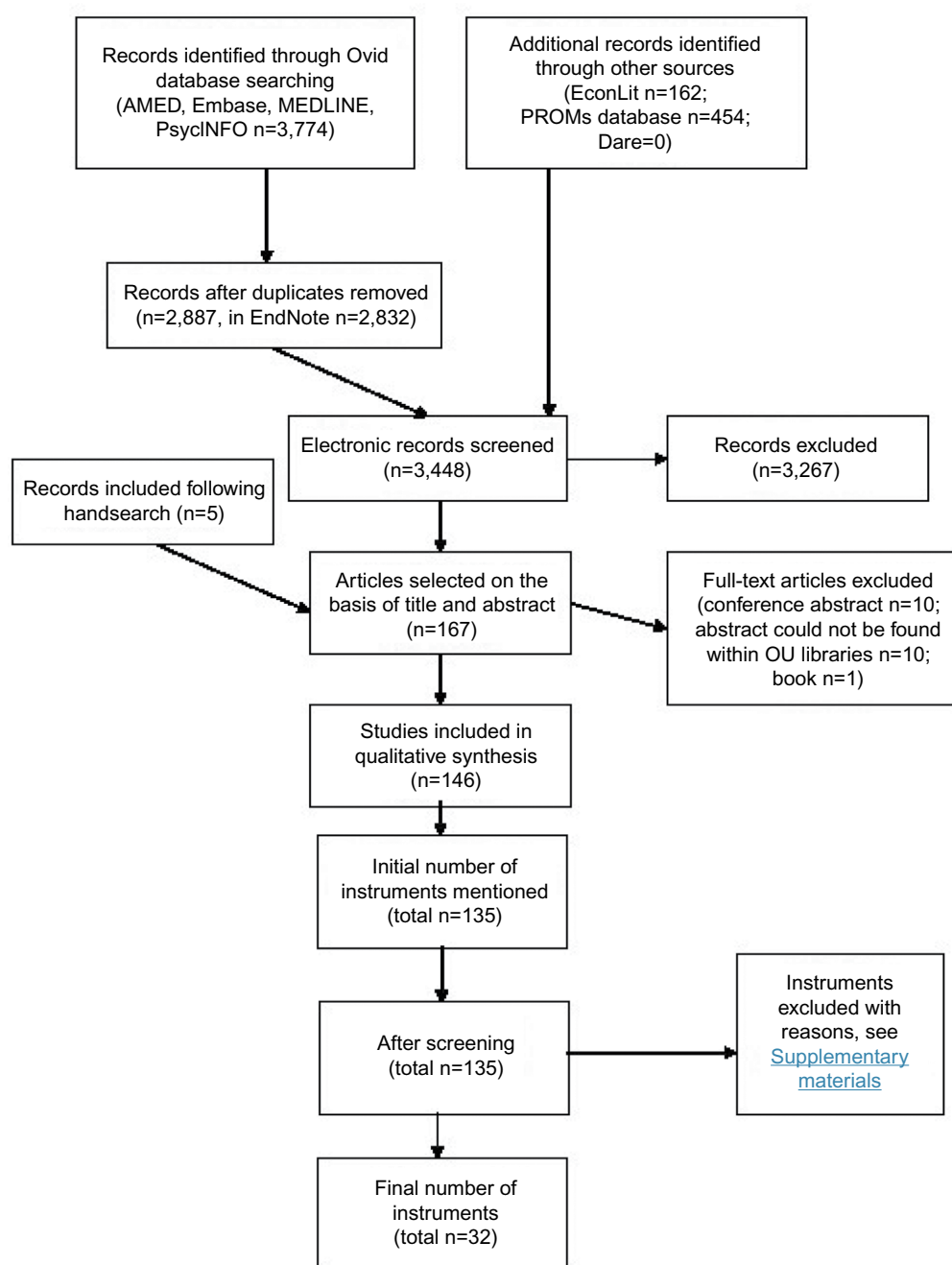


Figure 1 Instrument flow diagram.

Abbreviations: PROMs, patient-reported outcome measures; OU, Oxford University; AMED, Allied and Complementary Medicine Database; PROMs, patient-reported outcome measures.

hundred and thirty-five instruments were initially identified from the selected full-text articles. A reliability exercise was performed on 16 full-text articles between two reviewers, and the agreement was 95% (38/40 questionnaires identified). After screening, 67 instruments were left. Additionally, if the instrument was not validated (developed for or subsequently validated) to be used in a population of patients undergoing hip or knee replacement surgery, it was also excluded.

An instrument-specific search was performed on each of the 67 identified instruments. By this method, 21 new

validation papers (in addition to 42 developmental papers) in the targeted population were identified. Furthermore, on closer examination of shortlisted instruments, 21 initially identified instruments were additionally excluded (reasons listed in [Supplementary material 3](#)).

Data extraction

Relevant data on the psychometric performance and operational characteristics were extracted for each PROM. The summary texts ([Supplementary material 4](#)) were sent to

corresponding authors from the developmental study of each respective PROM, and further information was added as a result of this exercise. The appraisal summaries are presented in Tables 1–4.

Table 1 summarizes the evidence of measurement and operational performance applying the adapted appraisal criteria for the hip PROMs identified in this review. On the basis of the volume and quality of evidence, the Oxford Hip Score (OHS) clearly has the best evidence of measurement properties within the hip-specific PROM category. Within the “knee scores” subgroup (Table 1), the Oxford Knee Score (OKS; with the OKS-Activity and Participation Questionnaire, or OKS-APQ) demonstrated best evidence of its measurement properties within the knee-specific PROM category. The Knee Injury and Outcomes Osteoarthritis Score (KOOS) and

the KOOS-Physical Function Short Form have some favorable evidence of their measurement properties, although in comparison with the OKS, the evidence is lacking and further evaluations are needed.

Table 2 summarizes the evidence of measurement and operational performance, by applying the adapted appraisal criteria to the lower limb and pain PROMs identified in these reviews. The best-performing lower limb measure for hip/knee patients is the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC), followed by the Lower Extremity Functional Scale. The WOMAC also performed best when applied to separate hip or knee groups. Satisfactory evidence of its measurements properties was generally lacking for all of the three identified pain measures (ICOAP, P4, McGill Pain-Short Form). ICOAP and McGill Pain-Short

Table 1 Hip and knee scores

Instrument group	Hip				Knee				
Instrument (groups tested)	HOOS (h)	HRQ (h)	PSI (h)	OHS (h)	Knee disorders subjective history (VAS) (k)	KOOS (k)	KOOS-PS (k)	OKS (k)	OKS-APQ (k)
Number of studies	5	1	4	20	1	3	2	23	1
Reproducibility	++	+	+	++	0	+	0	+++	+++
Internal consistency	+	0	0	++	0	0	+++	+++	+++
Validity: content	0	0	++	++	+	+	+	+++	+++
Construct	++	+	++	+++	+	+	++	+++	+++
Responsiveness	+	+	++	+++	0	0	++	+++	+++
Interpretability	0	0	0	+++	0	0	0	++	0
Floor/ceiling/precision	+	0	0	–/+	0	+	0	++	++
Acceptability	0	0	0	+++	–	0	0	+++	+++

Notes: Psychometric and operational criteria: “0” not reported; “–” no evidence in favor; “+” some limited evidence in favor; “++” some good evidence in favor; “+++” good evidence in favor; “+/-” mixed evidence.

Abbreviations: HOOS, Hip Disability and Osteoarthritis Outcome Score; HRQ, Hip Rating Questionnaire; PSI, patient-specific index; OHS, Oxford Hip Score; KOOS, Knee Injury and Outcomes Osteoarthritis Score; KOOS-PS, KOOS-Physical Function Short Form; OKS, Oxford Knee Score; OKS-APQ, OKS-Activity and Participation Questionnaire; h, hip; k, knee; VAS, visual analog scale.

Table 2 Lower limb and pain scores

Instrument group	Lower limb							Pain		
Instrument (group tested)	LEFS (h/k)	WOMAC (h/k)	WOMAC (h)	WOMAC (k)	WOMAC SF (h/k)	Lower limb core score (h/k)	MODEMS-HK/ (AAOS) hip and knee core score (h/k)	ICOAP (h/k)	P4 (h/k)	McGill pain-short form (h/k)
Number of studies	5	25	N/a	N/a	N/a	1	1	2	1	2
Reproducibility	+	++	++	+	0	0	0	+	0	++
Internal consistency	+	+	0	0	+	0	0	+	++	0
Validity: content	+	+	+	+	+	+	+	++	+	0
Construct	++	+++	+	++	++	0	+	+	+	+
Responsiveness	++	+++	++	++	+	0	++	–	0	–
Interpretability	+	++	++	++	0	0	0	0	0	0
Floor/ceiling/precision	0	–/+	–	0	0	0	++	0	0	0
Acceptability	0	++	+	+	0	0	+	0	0	0

Notes: Psychometric and operational criteria: “0” not reported; “–” no evidence in favor; “+” some limited evidence in favor; “++” some good evidence in favor; “+++” good evidence in favor; “+/-” mixed evidence.

Abbreviations: LEFS, Lower Extremity Functional Scale; WOMAC, Western Ontario and McMaster Universities Osteoarthritis Index; WOMAC-SF, WOMAC Short Form; MODEMS-HQ, Musculoskeletal Outcome Data Evaluation and Management System Hip and Knee Core Scale; AAOS, American Academy of Orthopedic Surgeons; ICOAP, Intermittent and Constant Osteoarthritis Pain Measure; h, hip; k, knee; N/a, not applicable.

Table 3 Utility and generic scores

Instrument group	Utility					Generic						
	SF6D (h)	HUI2 and HUI3 (h)	EQ-5D (h/k)	EQ-5D (h)	EQ-5D (k)	SF-36 (h/k)	SF-36 (h)	SF-36 (k)	SF-12 (h/k)	SF-12 (h)	SF-12 (k)	SIP (h)
Number of studies	1	4	9	N/a	N/a	14	N/a	N/a	3	N/a	N/a	2
Reproducibility	0	0	0	0	0	0	0	0	++	0	0	0
Internal consistency	0	0	N/a	N/a	N/a	0	0	—	0	0	0	0
Validity: content	0	0	0	0	0	0	0	0	0	0	0	+
Construct	0	++	+	0	+	+	0	+	0	0	+	+
Responsiveness	++	+	0	0	+	0	++	+	0	+	+	—
Interpretability	0	0	0	++	++	0	+	+	0	+	+	0
Floor/ceiling/precision	—	0	0	0	++	0	—	0	+++	0	0	—
Acceptability	0	0	0	0	++	0	0	0	0	0	0	0

Notes: Psychometric and operational criteria: “0” not reported; “—” no evidence in favor; “+” some limited evidence in favor; “++” some good evidence in favor; “+++” good evidence in favor.

Abbreviations: SF6D, Short Form 6D; HUI2 and HUI3, Health Utilities Index Mark 2 and Mark 3; EQ-5D, European Quality of Life Questionnaire 5D; SF-12, Short Form 12; SF-36, short form 36; SIP, sickness impact profile; h, hip; k, knee; N/a, not applicable.

Table 4 Other scores

Instrument	WHOQOL-BREF (h/k)	Aberdeen IAP (h/k)	Aberdeen IAP (modified) (h/k)	NEADL (h)	AQOL (h/k)	MSK functional limitations index (k)	HAQ (k)	MHAQ (h/k)	MHAQ (h)	K10 (h/k)
Number of studies	1	1	1	1	2	1	2	2	N/a	1
Reproducibility	0	0	0	++	0	0	0	0	0	0
Internal consistency	++	+	++	++	0	0	—	0	0	0
Validity: content	+	+	0	—	0	0	0	0	+	—
Construct	0	+	+	+	+	+	++	+	+	+
Responsiveness	+	0	0	—	++	0	—	—	+	—
Interpretability	0	0	0	0	0	0	0	0	0	0
Floor/ceiling/precision	++	+	0	+	0	0	+	++	0	++
Acceptability	0	0	0	0	0	+	0	0	0	0

Notes: Psychometric and operational criteria: “0” not reported; “—” no evidence in favor; “+” some limited evidence in favor; “++” some good evidence in favor; “+++” good evidence in favor.

Abbreviations: WHOQOL-BREF, World Health Organization Quality of Life Instrument; Aberdeen IAP, Aberdeen Impairment, Activity Limitation, and Participation Restriction; NEADL, Nottingham Extended Activities of Daily Living; AQOL, Assessment of Quality of Life; MSK, musculoskeletal; HAQ, Health Assessment Questionnaire; MHAQ, Modified Health Assessment Questionnaire; K10, The Kessler Psychological Distress Scale; h, hip; k, knee; N/a, not applicable.

Form had no evidence in favor of their responsiveness and P4 did not have any reported evidence of its responsiveness. Three utility and generic measures identified in the review are listed in Table 3. As with the pain scores, the evidence for utility PROMs was generally lacking, with the European Quality of Life Questionnaire scoring worse on construct validity and responsiveness than the Short Form 6D, and Health Utilities Index Mark 2 and Mark 3 (HUI2 and HUI3). On the basis of the volume and quality of evidence, among all identified generic measures, the Short Form 12 (SF-12) is clearly the most promising one.

Nine measures identified in the review were categorized as “other” scales. Table 4 summarizes evidence of their measurement properties. World Health Organization Quality of Life Instrument (WHOQOL-BREF), Aberdeen Impairment, Activity Limitation, and Participation Restriction (Aberdeen

IAP), and Assessment of quality of life had the best overall evidence in this subcategory (on a mixed hip/knee population). However, the overall evidence of their validity was generally lacking.

Discussion

This review has examined, in great detail, measurement properties of PROMs used in patients undergoing hip and/or knee replacement surgery. Generally, sufficient information on measurement properties is lacking for a large number of PROMs. The best-performing site-specific PROMs were the OKS, OHS, and (lower limb specific) WOMAC. The best-performing generic measure was SF-12.

Alviar et al⁸ published a systematic review of measurement properties of 28 PROMs used in hip/knee arthroplasty based on published evidence up to December 2009 (although

using slightly different screening methods and appraisal criteria) and found WOMAC, OKS, and SF-36 to be the most comprehensively tested measures to date, with a need for more rigorous evaluation of reliability responsiveness and interpretability. Our review has updated this evidence, both in breadth (we have assessed 67 instruments) and time period (our search was until May 2014). Furthermore, a comprehensive review by Browne et al¹² favored the OHS and OKS (used alongside European Quality of Life Questionnaire 5D) as primary outcome measures of choice to be used in a national audit of hip and knee replacement surgery (NHS PROMS program).

The comprehensive evidence presented in this manuscript can be used by researchers, clinicians, and commissioners as a reference point when evaluating the potential usefulness of a measure in patients undergoing hip and knee arthroplasty, or informing the need for further research.

It should be noted that the standards (and indeed scope/tolerance) for reporting details of qualitative procedures and psychometric analysis have changed over the last 20 years, (very much so in the musculoskeletal literature), so that while measures that were devised earlier in that period have had longer time in which to accrue evidence of their measurement properties, they can frequently lack relevant detail specifically in relation to the development of the instrument. In the majority of cases, more detailed reports on PROMs will have been reported in more recent publications. This is probably in part a consequence of the evolution of methods (and terminology) over time and proliferation of recommendations about minimum standards and guidance on its reporting by various authors and dedicated organizations (eg, Streiner and Norman¹³; COSMIN; and the US Food and Drug Administration).^{2,14} For these reasons, finding and appraising different measures require a comprehensive literature search. In addition, in some cases (and for the purpose of clarification or obtaining relevant unpublished information), it is helpful to contact authors involved in the original developmental studies, as we have done in this study.

It is the authors' opinion that the preference for a primary outcome measure in this population of patients should be given to disease-/site-specific score, rather than generic, to ensure better coverage of the construct of interest and better responsiveness. In this review, we have identified the WOMAC, OHS, and OKS (with OKS-APQ) to be the most promising measures. Further research, however, on some of the missing measurement properties in these measures is wanting. For the WOMAC, further evidence on ceiling/floor effect, content validity, and acceptability is required in

both hip and knee groups of patients. The OHS is currently lacking evidence on its ceiling/floor effects, and the OKS-APQ does not have detailed evidence of its interpretability published.

Conclusion

This review identified and reviewed the psychometric quality of 32 PROMs, which were developed and/or have been used in patients undergoing hip and/or knee replacement surgery. A large number of these measures lack essential evidence of their measurement properties in these groups of patients. On the basis of the review criteria, the measures with most complete evidence to date are the OHS (for patients undergoing hip replacement surgery) and the OKS, with OKS-APQ (for patients undergoing knee replacement surgery). While less specific, the WOMAC is the second most promising measure. Further research, as outlined in this summary, is required on almost all of the identified measures. Researchers can use the information presented in this review to inform further psychometric studies of the reviewed measures.

Acknowledgments

A copy of the OHS, OKS, and OKS-APQ questionnaires and permission to use these measures can be acquired from Isis Innovation Ltd, the technology transfer company of the University of Oxford via website: <http://www.isis-innovation.com/outcomes/index.html> or email: healthoutcomes@isis.ox.ac.uk. The research reported in this manuscript was supported by the NIHR Health Technology Assessment program grant award: "Introducing Standardized and Evidence Based Thresholds for Hip and Knee Replacement Surgery – The Arthroplasty Candidacy Help Engine (The ACHE tool)", under the grant number 11/63/01.

Disclosure

The authors report no conflicts of interest in this work.

References

1. Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. *Health Technol Assess.* 1998;2(14):1–74.
2. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol.* 2010;63(7):737–745.
3. US Department of Health and Human Services, Food and Drug Administration. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims. *Health Qual Life Outcomes.* 2009;4:79.

4. Patrick DL, Burke LB, Gwaltney CJ, et al. Content validity – establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO good research practices Task Force report: part 2 – assessing respondent understanding. *Value Health*. 2011;14(8):978–988.
5. Patrick DL, Burke LB, Gwaltney CJ, et al. Content validity – establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO good research practices task force report: part 1 – eliciting concepts for a new PRO instrument. *Value Health*. 2011;14(8):967–977.
6. Rothman M, Burke L, Erickson P, Leidy NK, Patrick DL, Petrie CD. Use of existing patient-reported outcome (PRO) instruments and their modification: the ISPOR good research practices for evaluating and documenting content validity for the use of existing instruments and their modification PRO task force report. *Value Health*. 2009;12(8):1075–1083.
7. Garratt A, Schmidt L, Mackintosh A, Fitzpatrick R. Quality of life measurement: bibliographic study of patient assessed health outcome measures. *BMJ*. 2002;324(7351):1417.
8. Alviar MJ, Olver J, Brand C, et al. Do patient-reported outcome measures in hip and knee arthroplasty rehabilitation have robust measurement attributes? A systematic review. *J Rehabil Med*. 2011;43(7):572–583.
9. Garratt A, Brealey S, Gillespie W. Patient-assessed health instruments for the knee: a structured review. *Rheumatology*. 2004;43(11):1414–1423.
10. Smith SC, Cano S, Lamping DL, et al. *Patient-Reported Outcome Measures (PROMS) for Routine Use in Treatment Centres: Recommendations Based on a Review of the Scientific Evidence*. Final report to the Department of Health. London, UK: London School of Hygiene & Tropical Medicine; 2005.
11. Terwee CB, Jansma EP, Riphagen II, de Vet HC. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Qual Life Res*. 2009;18(8):1115–1123.
12. Browne J, Lewsey L, Van Der Muelen J, Black, N. *Patient Reported Outcome Measures (PROMS) in Elective Surgery*. Report to the Department of Health. London, UK: London School of Hygiene & Tropical Medicine; 2007.
13. Streiner DL, Norman GR. *Health Measurement Scales: A Practical Guide to Their Development and Use*. Oxford, UK: Oxford University Press; 2008.
14. Food and Drug Administration. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims. *Federal Reg*. 2009;74(235):65132–65133.

Patient Related Outcome Measures

Publish your work in this journal

Patient Related Outcome Measures is an international, peer-reviewed, open access journal focusing on treatment outcomes specifically relevant to patients. All aspects of patient care are addressed within the journal and practitioners from all disciplines are invited to submit their work as well as healthcare researchers and patient support groups.

Submit your manuscript here: <http://www.dovepress.com/patient-related-outcome-measures-journal>

Dovepress

The journal is included in PubMed. The manuscript management system is completely online and includes a very quick and fair peer-review system. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.